Evolving agents for personalized information filtering

Sheth, B. Maes, P.
MIT Media Lab., Cambridge, MA, USA ;

This paper appears in: **Artificial Intelligence for Applications, 1993. Proceedings., Ninth Conference on**

Meeting Date: 03/01/1993 - 03/05/1993
Publication Date: 1-5 March 1993
Location: Orlando, FL USA
On page(s): 345 - 352
Reference Cited: 14
Inspec Accession Number: 4857590

Abstract:
Describes how techniques from artificial life can be used to evolve a population of personalized information filtering agents. The technique of artificial evolution and the technique of learning from feedback are combined to develop a semi-automated information filtering system which dynamically adapts to the changing interests of the user. Results of a set of experiments are presented in which a small population of information filtering agents was evolved to make a personalized selection of news article from the USENET newsgroups. The results show that the artificial evolution component o the system is responsible for improving the recall rate of the selected set of articles, while learning from feedback component improves the precision rate

Index Terms:
feedback genetic algorithms information retrieval learning (artificial intelligence) online front-ends personal computing software agents USENET newsgroups artificial evolution artificial life changing user interests dynamic adaptation evolving agents learning from feedback news articles personalized information filtering precision rate recall rate semi-autonomous system

Documents that cite this document
Select link to view other documents in the database that cite this one.

Evolving Agents For Personalized Information Filtering

Beerud Sheth
MIT Media Lab
20 Ames St.
Cambridge, MA 02139
beerud@media.mit.edu

Pattie Maes
MIT Media Lab
20 Ames St.
Cambridge, MA 02139
pattie@media.mit.edu

Abstract

This paper describes how techniques from Artificial Life can be used to evolve a population of personalized information filtering agents. The technique of *artificial evolution* and the technique of *learning from feedback* are combined to develop a semi-automated information filtering system which dynamically adapts to the changing interests of the user. We present results of a set of experiments in which a small population of information filtering agents was evolved to make a personalized selection of USENET netnews messages for a particular user. The results show that the artificial evolution component of the system is responsible for improving the *recall rate* of the selected set of articles, while learning from feedback component improves the *precision rate*.

1 Introduction

One of the main problems in building a system for personalized information filtering is the construction of a profile of the user's information interests. There are three subproblems involved. The first is finding a representation for the user profile that allows both power and flexibility. Second, it is important that the user be able to communicate her desires and interests to the system so that an initial profile can be constructed. Finally, the system has to be responsive and change this initial profile as the interests of the user change over time.

This paper proposes to use techniques from the field of Artificial Life to build a personalized information filtering system. Artificial evolution – often implemented as a genetic algorithm – has proven to be an effective parallel search technique in a number of problem domains [2, 7, 8, 9]. It has also been shown that combining artificial evolution with individual learning by the evolved organisms speeds up the search process significantly [1, 10]. This combination of techniques is also particularly useful in situations where the optimal solution keeps changing over time. This property makes them attractive as a technique for searching the space of user profiles in an adaptive information filtering system.

The first section of this paper discusses the Information Filtering problem and presents a short overview of previous work in the field. We then show how a genetic algorithm combined with individual learning can be used for the search of a user profile.

Examples are presented from an implemented prototype which filters news articles from the USENET newsgroups. Experimental results demonstrate how the different mechanisms of the system relate to performance evaluation parameters. In particular, the results show that the technique of genetic variation is responsible for improving the *recall* of the set of articles retrieved, while the technique of learning from feedback is responsible for improving its *precision*. The last section presents some concluding remarks along with a discussion of future research.

2 Information Filtering

Information filtering has been used to describe a variety of processes involving the delivery of information to users. While information filtering is related to processes such as retrieval, routing, categorization, and extraction, the distinction needs to be made clear so as to focus on the specific research issues associated with filtering [3].

Information filters are mediators between sources of information and their end-users. Filtering applications typically involve streams of incoming data, either being broadcast by remote sources or sent directly by other sources. These data may also be the result of database searches. Information filtering is typically concerned with repeated uses of the system, by a person or persons with long-term goals or interests, unlike a typical information retrieval system.

Filtering mainly deals with a dynamic data stream, as opposed to a static database, from which texts are selected or eliminated. This also has a bearing on the performance evaluation criteria to be used for a filtering system. The user's mode of interaction with a filtering system is fairly different from other information gathering systems. Instead of responding to user interaction in a single information-seeking episode, a filtering system has to deal with long-term changes over a series of information-seeking episodes. Information filters are more likely to be personalised to serve the same user's need over a relatively long period of time. Learning and adaptation are, therefore, issues of prime importance to filtering systems [3].

Some of the research carried out in information retrieval is directly relevant to information filtering systems. Especially, work done in the areas of text representation, retrieval techniques, and user modelling can be leveraged to design better filtering systems. Con-

1043-0989/93 $03.00 © 1993 IEEE

ventional text representation schemes commonly use indexing methods, while more sophisticated schemes use clustering, boolean probabilistic models or vector spaces to represent texts [11]. Retrieval techniques are concerned with estimating the "score" of an object to be retrieved. Research in user modelling has been mainly focussed on query formulation and relevance feedback as mechanisms for the system to acquire information about the goals of the user. Performance of retrieval systems is shown to be significantly improved by using simple relevance feedback techniques [12].

A number of different approaches have been used to automate information filtering. Rule based systems which observe user's usage patterns and make suggestions based on them have been described in [4]. Rules are used to measure usage patterns such as commonly occuring terms, as well as timeliness measurements like frequency and recency. This helps in bringing usage patterns to the attention of the users. Statistical methods have been useful in improving filtering methods. [5] presents results of an experiment aimed at determining the effectiveness of four statistical information filtering methods in the domain of technical reports. A novel mechanism for collaborative filtering in which users annotate documents is presented in [6]. When new documents arrive, "eager readers" annotate the documents, while "casual readers" can install filters which use these annotations in addition to the content of the document.

One of the desirable features in an information filtering system is that they recommend new information not already in the profile, which might possibly be of interest to the user. A rule based system which looks for usage patterns can only comment upon what the user is already doing, not change it. One of the advantages of the artificial evolution approach described in this paper is its exploratory behavior. By mutations and crossovers of fit information filtering agents, the system can explore newer domains which may be of potential interest to the user. Another desirable feature is that the filtering system should be able to unlearn previously learned knowledge when the users interests change. A statistical profile builder might build a good user profile, but then there is a high inertia towards unlearning when necessary. In artificial evolution, agents have to continually gain fitness in succeeding generations, else they are eliminated from the population. This means that an agent which had a high fitness value in the last generation might not be able to survive to the next, if it does not gain fitness in the present generation. This enables the system to be dynamically adaptive to the user's interests.

3 The Algorithm

The problem of building a personalized Information Filtering system can be viewed as a search process. It involves searching over the large and complex space of possible user profiles, for an "optimal" user profile (or a set of profiles) that match the user's different interests. This "optimal" user profile has to vary as the user's interests change over time.

Evolution can be viewed as search in a space of genotypes for the ones that are the fittest (or the best adapted) to survive in the environment. Cycles of genetic variation followed by selection of the fittest produce a relatively fitter species with every generation. Genetic Algorithms extract and generalize critical processes of evolution and use them to solve artificial search problems [9]. They have proven very successful in searching for global optima in large and complex search spaces.

Searching a large and changing space involves a trade-off between two objectives: (*i*) *exploiting* the currently available solution and (*ii*) further *exploring* the search space for a possibly better solution. Hill Climbing is an example of a search technique which exploits the best known alternative. However, because of this very reason, it is likely to get stuck in local maxima. Random Search, on the other hand, is an extreme case of an exploring search technique: it is unsatisfactory as it does not make use of the best solution found so far. Genetic Algorithms manage the trade-off between exploration and exploitation in a near optimal way — they exploit the solution found so far, while Crossover and Mutation operations provide a way of exploring the search space for better solutions [9].

Several experiments have demonstrated that artificial evolution is helped by individual learning [1, 10]. This phenomenon is also known as the "Baldwin effect": if the organisms evolved are allowed to learn during their lifetime, then the evolution towards a fitter species happens much faster. This is the case because every individual is able to explore a "patch" of the search space (find the maximum fitness in the local neighbourhood of its genotype) rather than a single point (evaluate the fitness of its own genotype).

We have used a genetic algorithm with individual learning to build a prototype of a personalized Information Filtering system. Presently, we use USENET network news as the data stream from which articles are retrieved.

The system consists of a number of news categories which a user has defined [1]. Each of these news categories consists of a population of filtering agents. These are "organisms" that retrieve articles which match an internal representation of the type of article they are interested in. The internal representation consists of whatever the organism inherited genetically from its parents (the genotype) augmented with information it learns during its lifetime. Agents are assigned a fitness value based on the user feedback regarding their performance. The user conveys whether an article that was retrieved by one or several agents was appreciated or not. The agents learn from this feedback by changing their internal representation to reflect this training example. For each positive/negative feedback received, an agent gets positive/negative fitness points. To create the next generation of agents, only the very fit agents are selected to produce offspring. The offspring is produced by applying the copy, crossover and mutation operators to the fit agents.

This genetic process driven by user feedback makes

[1] The way in which a user can define a news category is explained later.

Xref: clari.news.economy:1925 clari.news.disaster:948
From: clarinews@clarinet.com
Newsgroups: clari.news.economy,clari.news.disaster
Subject: Prices decline on world markets despite hurricane
Keywords: oil, energy, economy, severe weather, trouble
Message-ID: <oilpriceU2aP5pc@clarinet.com>
Date: 25 Aug 92 22:09:12 GMT
References: <oilpriceU2l6540pe@clarinet.com>
Lines: 85
Approved: clarinews@clarinet.com
X-Supersedes: <oilpriceU2aP545pe@clarinet.com>
Location: texas
ACategory: national
Slugword: oilprice
Priority: major
Format: regular
ANPA: Wc: 868; Id: z6205; Sel: txbyo; Adate: 8-25-5pcd; Ver: 38/2
Codes: ybyortx., yne.rtx., ynbwrtx., xxxxxxxx

Figure 1: A sample news header of a "richer", more structured article.

the population of filtering agents to evolve towards the optimal interest profile of the user. The details of the algorithm are as described below.

3.1 Genotype and Internal Representation

Genotypes are the individual points in the search space of user profiles. A sample genotype is shown below[2]:

newsgroup:	*clari.sports.basketball*
location:	*boston, chicago, usa*
source:	*New York Times*
keywords:	*celtics, bulls, jordan, magic johnson*

At birth, an agent creates an *internal representation* based on its genotype. As the agent learns during its lifetime, changes are made to this internal representation. The internal representation is structured in the same way as the genotype described above. This way both the genetic algorithm as well as the learning from feedback mechanism search the same space of user profiles (which is necessary for the Baldwin effect to be able to take place).

The internal representation, when created, has exactly the same structure and information as the genotype. In addition it maintains weights for all of the attributes (such as keywords, source, authors, etc) as it learns that some are more relevant than others. The initial weights of the attributes are all small positive values. This ensures that, while the offspring inherits some attributes from the parents (a parental "bias"), attributes learned during the organism's lifetime also have a fair chance of proving their relevance.

[2] "location" and "source" are record fields provided by the Netnews database.

3.2 Learning from Feedback

When an agent receives positive feedback, it extracts information from the corresponding article and incorporates it into its internal representation. Presently, the agent extracts most of the information provided in the header of the news article (Figure 1), in particular the author, keywords, location, category and priority fields. If, say, a keyword is already present in the internal representation, it's weight is increased, so that the agent is more likely to retrieve similar articles in the future. Conversely, in the case of negative feedback, the information is stored with negative weight, so as to make it less likely that similar articles will be retrieved in the future.

The user can also manually indicate preference for particular keywords occuring in an article. This can be done by highlighting the appropriate words in the text of the article. These keywords (with initial small positive or negative weight) get added to the internal representation of the agent (if they already exist, their weights are increased or decreased respectively).

3.3 Phenotype

The phenotype is the manifested behavior of the agent in its environment. Each agent looks up the newsgroup as specified in its internal representation. Each article header is rated and assigned a relevance value. Relevance points are assigned to the article for each point of similarity to the internal representation. For example, for a keyword in the *subject* or *keywords* field of the article that matches one in the *keywords* field of the agent, points proportional to the weight of the keyword are assigned. The sum of all these relevance points determines the overall relevance score of the article. The articles with high scores are retrieved, the rest are filtered away. The number of articles retrieved by an agent for display to the user is proportional to the agent's fitness.

3.4 Fitness function

An agent (or phenotype) is assigned a fitness value based on the user feedback received on articles the user reads. For every article the user indicates liking or disliking[3], the agent(s) which were responsible for retrieving that article get positive or negative fitness points respectively. The interface mechanism for the user to indicate her preference is described in the following section.

3.5 Initial Population

The initial population of agents is created when the user creates a new news category. The user must specify the name of the news category and can also give additional keywords which will be added to the genotypes of the first generation of agents. Suppose the user creates the news category *sports*. The system then looks up the list of available newsgroups to find those which have sports articles (presently, it is just a keyword based search). If the number of these newsgroups is large enough to form a population (as specified by the parameter defining the population size dis-

[3] Ideally, we would like the system to be able to deduce this information automatically based on how much time the user spent reading the article in ratio to how long the article is.

cussed below), then newsgroups are randomly selected from this set and assigned to the new agents. The user specified keywords are assigned to the *keywords* field of the genotype.

Each of these newly created agents has identical fitness values. Starting out with an initial generation of agents consisting of randomly created agents constructed on the basis of the user's input, the system evolves several generations of agents (based on user feedback) which are gradually more focussed to those articles which the user likes.

3.6 Genetic Operators

The genetic operators employed to create new agents are the crossover and mutation operator. These operators are the driving force behind the search process of the genetic algorithm. The user can either explicitly indicate which agents in the current population should be used as the basis for mutation or crossover or, alternatively, the system can make this selection automatically based on the fitness of the different agents (the probability of an agent being selected to reproduce being proportional to its fitness). In addition to agents with new genotypes, the new generation will consist of copies of the most fit agents of the old generation.

The crossover operator exchanges the *newsgroup* fields of two parent agents to create two new offsprings *i.e.* one offspring inherits the *newsgroup* field from one parent, and the other fields from the other parent; and vice versa for the other offspring.

The mutation operator replaces the *newsgroup* field, with another randomly selected newsgroup. This newsgroup is selected from the set of newsgroups which are "similar" to the one being replaced. The set of similar newsgroups is found by looking for shared keywords in names of newsgroups. The similarity requirement is so that an offspring, while being distinct from the parent, should not be too different so as to take advantage of the traits learned by the parent [4].

To be more specific, the genetic operators actually refer to the internal representation when creating the offspring. This way the offspring does not only inherit genetic information from its parents, but also "learned" information. This simulates "cultural learning" in the population of agents (or offspring imitating the behavior learned by the parents). The way in which this is done is that only the attributes (e.g. author, keywords, etc) with high weights are inherited by the offspring. At the same time as retaining the best characteristics of the parent, the offspring is also open to newer influences, because the weights of the inherited keywords are reset to small positive values in the offspring.

3.7 GA Parameters

There are some parameters to the genetic algorithm such as population size, frequency of crossovers, probability of mutation, the number of news articles to be presented every day, etc. Some users might have stable, fixed interests regarding news articles and would prefer a low occurrence of mutations. For the moment these parameters have to be set by hand (default values are provided).

[4] In future work, we hope to implement other types of mutations, e.g. based on thesauri, etc.

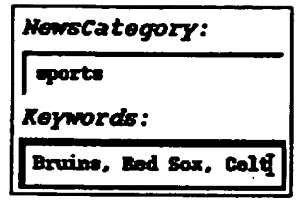

Figure 2: Creating a new news category

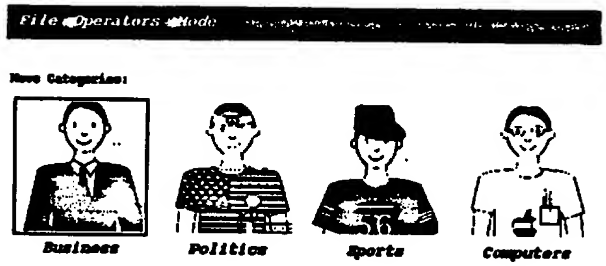

Figure 3: The news categories

4 User Interaction

One of the goals of this project is to make the user interaction as easy as possible. The user should be able to satisfy her goals with a minimal amount of interaction. In this section, we present a sample session which describes the way a user interacts with this system. This system was implemented in C++, Motif and BSD UNIX.

The user can define any number of news categories. A new news category can be created by specifying the name of the category and a set of keywords the user might be particularly interested in (as shown in Figure 2). The system then creates the initial population of agents for this news category, as described in the previous section.

Let's say the user has defined four categories, namely, *business*, *politics*, *sports* and *computers*. These categories are displayed to the user as shown in Figure 3 [5] The user can click on any of the icons to read the articles recommended by the agents in that news category. Figure 4 shows the articles selected for display by the agents in the *Business* news category. The articles selected by the agents in a population are all displayed together. Each of these articles is given

[5] The user can create her own icons.

a relevance score by the agent that selects it. The relevance score is displayed alongside the article title (as indicated by the number of stars prefixed to the article title). The user can see the contents of any of these selected articles by double-clicking on the appropriate article title.

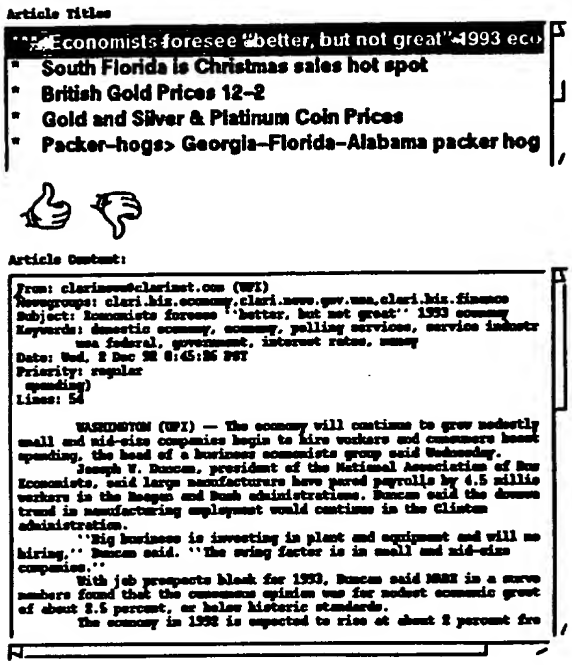

Figure 4: Article headers and body of one selected article within one news category. "Thumbs-up" and "thumbs-down" icons allow for user feedback.

The user can give positive or negative feedback by clicking on the "thumbs-up" or the "thumbs-down" icon respectively. Positive feedback for an article increases the fitness of the agent(s) which recommended it (and vice versa for negative feedback). The keywords, location, author and other information provided in the header of the article are incorporated into the internal representation of the agent. The user can also highlight a segment of text from the article body, and give positive feedback so that the selected text segment is included in the keywords field of the internal representation of the agent.

The interaction described above is the minimal amount of interaction a user needs to engage in to use this system. In the background, the system periodically creates new generations in which the good agents from the previous generation are retained, the unfit ones are "retired", and new agents are created using genetic operators on the fit agents. By just clicking on "thumbs-up" and "thumbs-down", the user is able to control the direction of evolution of populations of information filtering agents.

A more sophisticated user of the system might want to be able to exercise greater control over the population of agents. For example, the user can modify the survival threshold, the regeneration rate, the population size and other parameters which control the behavior of the population as a whole. This type of user might also want to go down to the level of individual agents and manipulate their internal representations, namely, the set of keywords, their weights, the newsgroup searched, etc. The system allows the user to have access at any of these levels [6] and be able to modify any component of the system.

5 Results

We have performed initial user tests of the system described above. Three different users who were not involved in the implementation of the system were asked to use the personal retrieval system during one whole week. They were also asked to use the regular USENET navigational interface to retrieve any articles they were interested in that were not retrieved automatically. All of their actions with both interfaces were recorded. This way, we were able to compare the set of articles retrieved automatically with the "optimal" set of articles (the set of articles that should have been retrieved). While a thorough analysis still remains to be done, these initial results have been encouraging.

Two main parameters of information retrieval effectiveness are *recall*, defined as the proportion of relevant articles retrieved, and *precision*, defined as the proportion of retrieved articles that are relevant [11]. While these parameters are not enough, in general, to completely evaluate the performance of an information filtering system, they are useful indicators.

Figure 5 contains three plots of recall and precision values with respect to the number of trials for three different users. To measure precision, the number of articles that were retrieved, and the number of articles read by the user were recorded in a manner transparent to the user. Precision was calculated as the percentage of retrieved articles that were read by the user. To find the articles that are relevant to the user, a simple interface to USENET was provided and users were asked to browse through the database and indicate the articles they would have liked the system to retrieve. This information was also recorded. Recall was calculated as the ratio of the retrieved articles read by the user over the union of those articles with the articles retrieved by hand by the user.

The graphs demonstrate that the recall as well as the precision of the set of articles retrieved improves

[6] A graphical interface for this level of interaction has yet to be built.

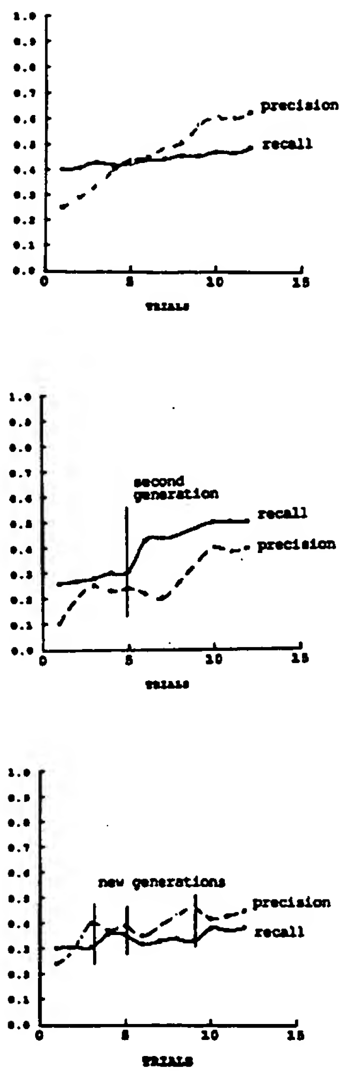

Figure 5: Precision and recall results for three different users.

the longer the system is in use. The first user (Figure 5a) did not use any genetic operators. There is an improvement in precision, however the recall value shows minimal changes. As the agents are rewarded for getting relevant articles, they get better at eliminating irrelevant articles. However, since there are no new newsgroup introduced through genetic operators, the recall rate does not improve much.

In the case of the second user (Figure 5b), a second generation of agents is created after 5 trials by applying genetic operators to the successful half of the population of agents. This adds new newsgroups, which helps improve the recall. There is a slight decrease in precision, because the offspring has lost some of the information learned by the parents during their lifetime. This decrease is not too significant, because the offspring inherits the fittest attributes from the parents.

The third user (Figure 5c) applied genetic operators more frequently. In some cases, the newly added newsgroups cause a decrease in recall as there is an inherent element of randomness. However, repeated negative feedback decreases the fitness of these undesirable newsgroups which are then eliminated when genetic operators are applied the next time.

In any automatic information retrieval system there is always a tradeoff between precision and recall (when both variables already have fairly good values). If one improves recall, then typically precision becomes worse and vice versa. One of the advantages of a genetic approach is that the user can dictate his/her own preferred trade-off of recall and precision by controlling the frequency with which genetic operators and feedback are applied. In further research we hope to demonstrate that if the agents are developed properly, it is also possible that high values of both precision and recall can be achieved simultaneously.

These results can be better understood with the help of a schematic diagram. In Figure 6, the circle represents the set of all articles. The region representing the set of relevant articles is shaded by vertical dotted lines. The articles retrieved by the filtering system are represented by horizontal dotted lines. For narrow or focussed filters, the precision is high — almost everything retrieved is relevant — but the recall is low since very few articles are actually retrieved (Figure 6a). As the search is broadened, the total number of relevant items retrieved goes up, enhancing the recall; at the same time, the number of nonrelevant retrieved items also grows, decreasing the precision (Figure 6b). That is, narrow searches produce high precision and low recall, whereas broader searches produce the reverse result [11].

The results obtained in our experiments suggest that in using genetic algorithms, learning from feedback helps improve precision. In terms of the user profile, this is a specialization of user's interests. Initially, the search is too broad; hence, the recall is moderately high while the precision is low. However, as the agents get feedback for the articles which the user thinks are relevant (or irrelevant), it retrieves fewer irrelevant articles. Hence the precision keeps improving over time without affecting recall too much

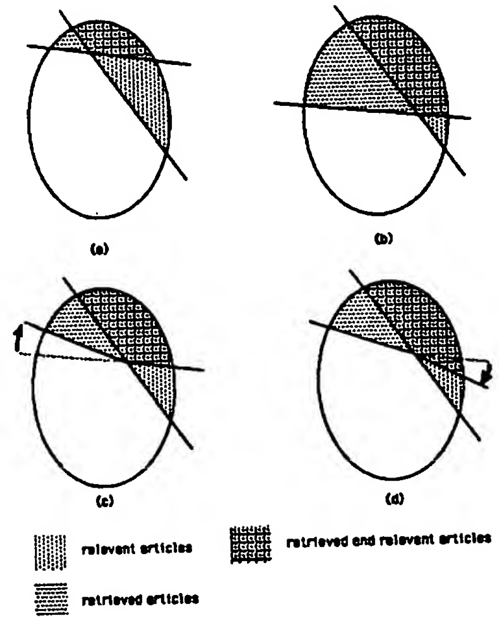

Figure 6: Recall and precision of the set of articles retrieved in the case of a) narrow search and b) broad search. c) The effects of learning on the set of articles retrieved : specialization. d) The effects of genetic operators on the set of articles improved : exploration.

(Figure 6c).

Genetic operators on the other hand are responsible for increasing the recall without sacrificing too much precision (Figure 6d). This corresponds to **exploration** of areas which may be of potential interest to the user. Mutation introduces a random newsgroup that had not been considered before and which the user might find relevant. This helps to retrieve proportionally more relevant articles, and thereby increases recall. At the same time, the mutated offspring is quite similar to the parent — the inherited precision not much worse than that of the parent genotype since the weakest gene was mutated. In case of crossover, the offspring retains the best features of the parents, thereby retaining most of the precision learned by the parents. At the same time, it also introduces newer kinds of articles which the user might possibly like, so as to help the recall.

In all of the cases studied, the users experienced a reduction in time and effort it takes to read news on a daily basis. We will have to test the system for longer periods of time to find out whether after a while the precision and recall rate approach numbers that make it acceptable to have a purely automated system (as opposed to a combination of manual and automated selection). The system would have been much more efficient if the various news databases had provisions for more feature descriptions of articles. In some newsgroups this is already the case. For example, the article header in Figure 1 contains various features such as keywords, location, category and priority. We expect the system to improve a lot once such additional features are taken into account.

6 Conclusion and Future Directions

The paper demonstrated that techniques from Artificial Life, in particular a combination of a Genetic Algorithm with Learning from Feedback, can be used to evolve a personalized system for automatic information filtering. Because of its dynamic nature, this system is able to adapt to the changing interests of the user.

We discussed a first prototype which assists the user in retrieving USENET Netnews articles. Results obtained in experiments with this system indicate that the genetic algorithm is responsible for improving the recall rate of the articles retrieved, while the learning mechanism is responsible for improving the precision rate.

While the first prototype produced some promising results, a lot of future research needs to be performed. The internal representation of our retrieving agents can be much improved. We intend to research more sophisticated representations which can represent more complicated user interests. We further intend to elaborate the graphical aspects of the user interface so as to present the user with an animated, graphical world of information agents. Eventually, we plan to hand the system to users for longer periods of time so as to thoroughly evaluate the premises of the project.

Acknowledgements

The first author acknowledges the support provided by the MIT AI Laboratory in this work. We also thank Jeff Bilmes for his comments on the earlier drafts of this paper. Finally, we are grateful to Pushpinder Singh, Ravi Sundaram and Sanjay Noronha for providing valuable comments on the system.

References

[1] Ackley, D., Littman, M., Interactions between Learning and Evolution, *Artificial Life II*, Edited by C. Langton, C. Taylor, J. Farmer and S. Rasmussen, Addison Wesley, 1991.

[2] DeJong, K.A., Adaptive System Design: A Genetic Approach, *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 10 No. 9, 1980.

[3] Belkin, N.J., Croft, W.B., Information Filtering and Information Retrieval: Two Sides of the Same Coin?, *Communications of the ACM*, Vol. 35 No. 12, pp. 29-38, 1992.

[4] Fischer, G., Stevens, C., Information access in complex, poorly structured information spaces. *Human Factors in Computing Systems CHI'91 Conference Proceedings*, 1991, pp. 63-70.

[5] Foltz, P.W., Dumais, T., Personalized Information Delivery: An Analysis of Information Filtering Methods, *Communications of the ACM*, Vol. 35 No. 12, pp. 51-60.

[6] Goldberg, G., Nichols, D., Oki, B.M., Terry, D., Using Collaborative Filtering to Weave an Information Tapestry, *Communications of the ACM*, Vol. 35 No. 12, pp. 61-70.

[7] Grefenstette, J.J., *Proceedings of an International Conference on Genetic Algorithms and Their Applications*, The Robotics Institute of Carnegie Mellon University, Pittsburgh, 1985.

[8] Grefenstette, J.J., Optimization of Control Parameters for Genetic Algorithms, *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 16 No. 1, 1986

[9] Holland, J.H., Adaptation in Natural and Artificial Systems, *An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*, University of Michigan Press, Ann Arbor, 1975.

[10] Hinton, G.E., Nowlan, S.J., How Learning can Guide Evolution, *Complex Systems*, 1: 495-502, 1987.

[11] Salton, G., *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*, Addison Wesley Publishing, 1989.

[12] Salton, G., Buckley, C. Improving retrieval performance by relevance feedback. *JASIS 41*, 1990, pp. 288-297.

[13] Salton, G., and Crouch D.B., User-System Interaction in Automatic Information Retrieval, TR 89-999, Department of Computer Science, Cornell University, Ithaca, 1989.

[14] Schank, Roger C., and Michael Lebowitz, The Use of Stereotype Information in the Comprehension of Noun Phrases, Alexandria, VA: Defense Technical Information Center, 1979.

Lawrence Hunter

UCHSC, Box C236
4200 E. 9^{th} Ave.
Denver, CO 80262
(303) 315-1094
FAX: (303) 315-1098
email: Larry.Hunter@uchsc.edu

180 Glencoe St.
Denver, CO 80220
(303) 320-6340
FAX: (303) 320-6879
hunter@180glencoe.com

Research Interests:

Development and application of advanced computational techniques for biomedicine, particularly the application of machine learning and statistical inference techniques to high-throughput molecular assays. Also, automated processing of biomedical texts, anatomically realistic models of neural computation, and neurobiologically and evolutionarily informed computational models of cognition.

Education:

B.A. in Psychology, 1982, Yale University, *cum laude.*
M.S. and M.Phil. in Computer Science, 1987, Yale University.
Ph.D. in Computer Science, 1989, Yale University.
Thesis: *Knowledge Acquisition Planning: Gaining Expertise Through Experience,* advised by Roger Schank.

Experience:

Associate Professor, 2000-:
University of Colorado School of Medicine, Department of Pharmacology
University of Colorado School of Medicine, Department of Preventive Medicine and Biometrics
University of Colorado, Boulder, Department of Computer Science
University of Colorado, Denver, Department of Biology

Molecular Mining Corp., Founder and member of the Board of Directors, 1997-2003.

Consultant, 1997-. Advise pharmaceutical and other biomedical industry clients on the applications of machine learning to problems in drug discovery, health care finance and other areas.

Freelance writer, 1987-. Articles on machine learning, privacy, biotechnology and social issues involving technology for academic, popular and industrial audiences.

National Cancer Institute, chief of section on Molecular Statistics and Bioinformatics, 1999-2000. Conduct basic research and supervise a team of M.S. and Ph.D. researchers in computational biology and machine learning; provide postdoctoral training; serve on NCI committees on bioinformatics.

George Mason University, Adjunct Associate Professor, 1991-2000. Teach graduate courses on computational biology in the Computational Science and Informatics program, and advise PhD student theses.

Krasnow Institute of Advanced Study in Cognition, Fellow, 1995-2000.

National Library of Medicine, Computer Scientist, 1989-1999. Director, Machine Learning Project. Responsible for conducting basic research on machine learning in biomedical domains. Project officer on AI software development contracts. Supervised medical students in the NLM Medical Informatics Elective.

Yale University, Instructor & Teaching Assistant, 1983-1988. Graduate course in knowledge representation and

memory, and undergraduate courses in artificial intelligence and computer programming.

Honors and Awards:

Engelmore Prize for Innovative Applications of Artificial Intelligence, 2003 (American Association for Artificial Intelligence)

Fellow, American College of Medical Informatics, 2002-

Regent's Award for Scholarship and Technical Achievement, (the highest honor granted by the National Library of Medicine), 1994.

Meritorious Service Award, National Library of Medicine, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998

Winner, student paper competition, Knowledge Acquisition for Knowledge Based Systems Workshop, 1988.

National Merit Scholar, 1978

Professional Activities:

Director, Center for Computational Pharmacology, 2000-

Advisory Board, University of Colorado Center for Computational Biology, 2001-

Board of Directors, International Society for Computational Biology, 1996-

Member of the University of Colorado Biomolecular Structure Program, 2001-

Member of the University of Colorado Cardiovascular Institute, 2001-

Member of the University of Colorado Cancer Center, 2000-

Member of the University of Colorado Human Medical Genetics, 2001-

Associate Editor, Journal of Biomedical Informatics, 2002-

National Academy of Sciences ad hoc reviewer, 2003

Study section, Bioinformatics Training Grant Review (NLM) 2001

President, International Society for Computational Biology, 1996-2000

Grant review study section, Neuroinformatics (NIMH) 2000

Grant review study section, Human Brain Project, 1997, 1998, 1999

Contract review study section, National Institute of Mental Health, 1998.

Cooperative Research and Development Agreement (CRADA) with VIPS Systems, Inc. 1998-2000.

Board of directors, National Science Foundation Scientific Database Network Project, 1992-1996

Board of directors, International Biomatrix Society, 1991-1996

Associate Editor, *Journal of Artificial Intelligence Research,* 1993-1997

Editorial board, *Artificial Intelligence and Medicine*, 1993-1995

Editorial board, *Journal of Computational Molecular Cell Biology,* 1993-1998

Special editor, *IEEE Expert*, track on Molecular Biology Applications, 1996.

Chapter chair (Washington, DC), Computer Professionals for Social Responsibility, 1992-1997

Program co-chair, First International Conference on Intelligent Systems in Molecular Biology, 1993

Organizing Committee, International Conference on Intelligent Systems in Molecular Biology, 1995, 1996

Program chair, Biotechnology Computing Track, Hawaiian International Conference on System Sciences, 1993, 1994, 1995.

Co-chair, Pacific Symposium on Biocomputing, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003

Area Chair for Machine Learning, National Conference on Artificial Intelligence, 1992, 1993.

Program chair, Biotechnology Computing Minitrack, Hawaiian International Conference on System Sciences , 1992

Program committee, AAAI-91 Workshop on Pattern Recognition and Inference in Molecular Biology, 1991

Program committee, International Machine Learning Conference, 1991

Site visit team, Human Genome Program Center Grants, 1991

Program chair, Biotechnology Computing Minitrack, Hawaiian International Conference on System Sciences - 24, 1991

Publications committee, American Medical Informatics Association, 1990-1993

Program chair, AAAI Spring Symposium on Artificial Intelligence and Molecular Biology, 1990.

Program committee, International Conference on the Biomatrix, 1990

NIOSH Safety and Occupational Health Study Section, 1989-1990

Books:
Editor, *Artificial Intelligence and Molecular Biology*, AAAI/MIT Press, 1993. Now available free online as http://www.aaai.org//Library/Books/Hunter/hunter.html

The Processes of Life, forthcoming from MIT Press.

Patents:
A System for Synergistic Combination of Multiple Automatic Induction Methods and Re-Representations of Data. US Patent 6,449,603, issued September 10, 2002. Licensed to firms in healthcare, insurance and the pharmaceutical industry.

Ph.D. Dissertations Directed:

Jeffery L. Krichmar, *A Computational Model of Cerebellar of Saccadic Control,* GMU Computational Science and Informatics, 1997.

Judith E. Devany, *Equation Discovery Through Global Self-Referential Geometric Invariants and Machine Learning,* GMU Information Technology, 1997.

Imran Shah, *Predicting Enzyme Function from Sequence,* GMU Computational Science and Informatics, 1998

Barry Zeeberg, *Whole Genome Information Analysis and Processing*, GMU Computational Science and Informatics, 1999

Robert S. Erb, *Analysis and Modeling of Gene Expression Circuits*, GMU Computational Science and Informatics, 1999

Myriam Abramson, *Learning Coordination Strategies* GMU Information Technology, 2003.

Lorraine Tanabe, *Text mining the biomedical literature for genetic interactions* GMU Computational Science and Informatics, 2003

Ronald Taylor *Reconstruction of metabolic and genetic networks from gene expression perturbation data using a Boolean model: construction of a simulation testbed and an empirical exploration of some of the limits* GMU

Computational Science and Informatics, 2003.

Peer Reviewed Publications:

Hu, X., Friedman, D., Hill, S., Caprioli, R., Powers, A, **Hunter,** L, and Limbird, L. Proteomic exploration of pancreatic islets from wildtype mice and mice lacking expression of the insulin release – suppressing α_{2A} adrenergic receptor. Submitted to *Molecular Pharmacology*, 2003

Hunter, L. *Life and Its Molecules: A Brief Introduction,* AI Magazine, to appear spring 2004

Rudolph,.M.C., McManaman, J.L., **Hunter, L.**, Phang,T., Neville,M.C. Initiation of Lactation in the Murine Mammary Gland: Temporal analysis of a complex biological switch with expression profiling and trajectory clustering. *Journal of Mammary Gland Biology and Neoplasia*, to appear, 2003.

Witzmann, F.A., Li, J., Strother, W.N., McBride, W.J., **Hunter, L.**, Crabb, D.W., Lumeng, L., Li, T.K. Innate Differences in Protein Expression in the Nucleus Accumbens and Hippocampus of Inbred Alcohol-Preferring (iP) and -Nonpreferring (iNP) Rats. *Proteomics* 2003 Jul;3(7):1335-44

Phang, T.L, Neville, M.C., Rudolph, M. and **Hunter, L.** Trajectory clustering: A non-parametric method for grouping gene expression time courses, with applications to mammary development., *Pacific Symposium on Biocomputing* 2003, 8:351-362.

Shenkar, R., Elliott, J.P., Diener, K., Gault, J., Hu, L.J., Cohrs, R.J., Phang, T., **Hunter, L.**, Breeze, R.E., and Awad, I.A., Gene Expression in Human Cerebral Vascular Malformations , *Neurosurgery, 2003 52(2):465-478*

Hunter, L. Ontologies for Programs, Not People. *Genome Biology* 2002, 3(6):interactions1002.1-1002.2

Cohen, K.B., Dolbey, A., Acquaah-Mensah, G. and **Hunter, L.** Contrast and variability in gene names *Proceedings of the Workshop on Natural Language Processing in the Biomedical Domain*, Philadelphia, July 2002, pp. 14-20 Association for Computational Linguistics.

Edgerton, ME, Taylor, R., Powell, JI., **Hunter, L.**, Simon, R., and Liu, E., A Bioinformatics Tool to Mine Sequences for Microarray Studies of Mouse Models of Oncogenesis, *Bioinformatics*, 18(5):774-775. 2002

Hunter, L., Taylor, R., Leach, S., & Simon, R., GEST: A Gene Expression Search Tool Based on a Novel Bayesian Similarity Metric, *Bioinformatics*. 2001 Jun;17 Suppl 1:S115-S122.

Shah, I. & **Hunter, L.** Visual Management of Large Scale Data Mining Projects,. *Pacific Symp. on Biocomputing*, 5:275-287, 2000

Tanabe, L., Rindflesch, T.C., Weinstein, J.N., **Hunter, L.**, Edgar: Extraction of Drugs, Genes and Relations from the Biomedical Literature, *Pacific Symposium on Biocomputing*, 5:514-525, 2000

Tanabe L, Scherf U, Smith LH, Lee JK, **Hunter L**, Weinstein JN., MedMiner: an Internet text-mining tool for biomedical information, with application to gene expression profiling. *Biotechniques*. 1999 Dec;27(6):1210-4, 1216-7.

Shah, I. & **Hunter, L.** Identification of divergent functions in homologous proteins by induction over conserved modules. *Intelligent Systems for Molecular Biology* 6:157-64 (1998)

Shah, I. & **Hunter, L.** Visualization Based on the Enzyme Commission Nomenclature. *Pacific Symposium on Biocomputing* 3:142-152 (1998).

Zeeberg, B.R. & **Hunter, L.** Characterization of a Family of Chimeric Proteins, the Amino Acyl tRNA Synthetases, by Determining Differential Codon Usage using One and Two State HMMs. *Mathematical Modeling and Scientific Computation,* 9(1):58-67, 1998.

Zeeberg, B.R. & **Hunter,** L. A Hidden Markov Model Whose Alphabet Is Nucleic Acid Triplet Codons and its Use

to Discover Chimerism in Protein Families, *Intelligent Systems for Molecular Biology* 5:153-156 , Menlo Park, CA: AAAI Press, 1997

Shah, I. & **Hunter, L.** Functional Classification of Enzymes by Sequence Alignment, *Intelligent Systems for Molecular Biology,* 5:276-83 , Menlo Park, CA: AAAI Press 1997

Krichmar, JL, Ascoli, G.A., Olds, J.L. and **Hunter, L.** A model of cerebellar saccadic motor learning using qualitative reasoning, *Biological and Artificial Computation: From Neuroscience to Technology* 1240: 133-145 (1997)

Krichmar, JL, Olds, JL. & **Hunter, L.** Qualitative Neurobiology, *Proceedings of the 1997 Workshop on Qualitative Reasoning,* pp. 265-276,1997

Abramson. M. Z. and **Hunter, L.**. Classification using Cultural Coevolution and Genetic Programmin. *Genetic Programming: Proc. of the First Anuual Conf.* 1996, pp. 249-254, MIT Press, 1996

Hunter, L. Coevolution Learning: Synergistic Evolution of Learning Agents and Problem Representations, *Proceedings of 1996 Multistrategy Learning Conference*, pp. 85-94, Menlo Park, CA: AAAI Press, 1996.

Dowe, D., Allison, L., Dix, T., **Hunter, L.**, Wallace, CS., & Edgoose, T., Circular Clustering of Protein Dihedral Angles by Minimum Message Length, *Pacific Symposium on Biocomputing* (1):242-255. World Scientific Press, 1996.

Harris, N., **Hunter, L.** & States, DJ. ClassX: A Tool for Browsing Protein Sequence Megaclassifications, *Proceedings of the Twenty-Sixth Annual Hawaii International Conference on System Sciences,* vol. 1, Los Alamitos, CA: IEEE Computer Society Press, Jan 1993; pp 554-563.

Hunter, L. AI and Grand Challenges in Biotechnology Computing, *Proceedings of the 13th International Joint Conference on Artificial Intelligence,* Morgan Kaufman, San Mateo, CA, Vol. 2, pp. 1677-1683, 1993.

Hunter, L. & Ram, A., Goals for Learning and Understanding. *Journal of Applied Intelligence.* 2(1):47-73, 1992.

Hunter, L. Knowledge Acquisition Planning: Using Multiple Sources of Knowledge to Answer Questions in Biomedicine, *Mathematical and Computer Modeling,* 16(6/7):79-91, 1992.

Hunter, L. & States, DJ., Bayesian Classification of Protein Structure, *IEEE Expert,* 7(4):67-75, 1992.

Hunter, L., Harris, N. & States, DJ. Efficient Classification of Massive, Unsegmented Datastreams, *Proceedings of the Ninth International Workshop on Machine Learning*, pp. 224-233, 1992, Morgan Kaufmann Associates, San Mateo, CA.

Hunter, L., Harris, N. & States, DJ. Megaclassification: Discovering Motifs in Massive Datastreams, *Proceedings of the Tenth National Conference on Artificial Intelligence,* pp. 837-842, 1992, AAAI Press, Menlo Park, CA.

Hunter, L. Bayesian Classification of Protein Structure Fragments, *The Proceedings of The Twenty Fourth Annual Hawaii International Conference on System Sciences; vol. 1.* Los Alamitos, CA: IEEE Computer Society Press. Jan. 1991; 595-604

Hunter, L. Artificial Intelligence and Molecular Biology, *AI Magazine* 11(5):27-36, 1991 Supplement.

Hunter, L. Applying Bayesian Classification to Protein Structure, *Proceedings of the Seventh Conference on Artificial Intelligence Applications,* vol. 1. Los Alamitos, CA: IEEE Computer Society Press. Feb. 1991; 10-16.

Hunter, L. & Ram, A. The Use of Explicit Goals for Knowledge to Guide Inference and Learning, *Proceedings of the Eighth International Workshop on Machine Learning,* Chicago, IL, June 1991, pp. 265-269, Morgan Kaufmann, San Mateo, CA.

Hunter, L. Knowledge Acquisition Planning for Inference from Large Datasets, *The Proceedings of The Twenty Third Annual Hawaii International Conference on System Sciences, Kona, HI. vol. 2, Software track*, pp. 35-44.

IEEE Press, 1990.

Hunter, L. Planning to Learn, *The Proceedings of The Twelfth Annual Conference of the Cognitive Science Society*, Boston, MA., July 1990, pp. 26-34, Lawrence Erlbaum Associates, Hillsdale, NJ.

Hunter, L. *Estimating Human Cognitive Capacities Cognitive Science, 12(2):257-261, April-June 1988*

Hunter, L. Explanation Based Discovery. *Proceedings of the AAAI Symposium on Explanation Based Learning*, Stanford, CA, March 1988, pp. 2-7.

Hunter, L. Artificial Neural Networks as Theories of Mind. *Proceedings of First Annual Conference of the International Neural Network Society*, Boston MA, September, 1988, IEEE Computer Society Press, Los Alamitos, CA.

Hunter, L. Knowledge Acquisition Planning. *Third Knowledge Acquisition for Knowledge Based Systems Workshop*, Banff, Alberta, Canada, November, 1988

Hunter, L. and Silbert, J. Progress Report on IVY: A Learning System for Information Retrieval in Pathology, . *Proceedings of the Artificial Intelligence and Medicine Workshop,* Seattle WA, 1987.

Collins, G., **Hunter,** L. and Schank, R. Transcending Inductive Category Formation in Learning, *Behavioral and Brain Sciences*, 9(4):639-686, December 1986.

Hunter, L. Steps Toward building a Dynamic Memory. *Proceedings of the Third International Workshop in Machine Learning*, Skytop, PA, June 1986, p.70-74, Morgan Kaufmann Associates, San Mateo, CA

Hunter, L. Indexing and Recognition: Metaknowledge for Organizing Information. *Proceedings of AI/BioMed: The First International Conference on Artificial Intelligence and its Impacts in Biology and Medicine,* Montpellier, France, September 1986, p.93-5

Book Chapters

A New Personal Right for the Information Age, with James B. Rule, in *Visions for Privacy*, Collin Bennett and Rebecca Grant (eds.) University of Toronto Press, 1999.

The Qualitative Reasoning Neuron: A New Approach to Modeling in Computational Neuroscience, with Jeffrey L. Krichmar, Giorgio A. Ascoli, and James L. Olds. in *Computational Neuroscience*, James Bower (ed), Plenum Press, NY. 1998

Planning to Learn, in *Goal-Driven Learning,* ed. by Ashwin Ram and David B. Leake, MIT Press, 1995.

The Use of Explicit Goals for Knowledge to Guide Inference and Learning, with Ashwin Ram, in *Goal-Driven Learning,* ed. by Ashwin Ram and David B. Leake, MIT Press, 1995.

Classifying for Prediction: A Multistrategy Approach to Predicting Protein Structure, in *Machine Learning IV*, ed. by R. Michalski & G. Tegucci, Morgan Kaufmann, 1994.

Planning to Learn About Protein Structure, in *Artificial Intelligence and Molecular Biology,* L. Hunter, ed., AAAI Press, 1993.

An Introduction to Molecular Biology for the Computer Scientist, in *Artificial Intelligence and Molecular Biology,* L. Hunter, ed., AAAI Press, 1993.

Proceedings Edited

Editor, with Russ B. Altman,A. Keith Dunker and Teri E. Klein, *Pacific Symposium on Biocomputing '03,* Singapore: World Scientific Press, January 2003

Editor, with Russ B. Altman,A. Keith Dunker and Teri E. Klein, *Pacific Symposium on Biocomputing '02,*

Singapore: World Scientific Press, January 2002

Editor, with Russ B. Altman,A. Keith Dunker and Teri E. Klein, *Pacific Symposium on Biocomputing '01,* Singapore: World Scientific Press, January 2001

Editor, with Russ B. Altman,A. Keith Dunker and Teri E. Klein, *Pacific Symposium on Biocomputing '00,* Singapore: World Scientific Press, January 2000

Editor, with Russ B. Altman,A. Keith Dunker and Teri E. Klein, *Pacific Symposium on Biocomputing '99,* Singapore: World Scientific Press, January 1999.

Editor, with Russ B. Altman,A. Keith Dunker and Teri E. Klein, *Pacific Symposium on Biocomputing '98,* Singapore: World Scientific Press, January 1998.

Editor, with Russ B. Altman,A. Keith Dunker and Teri E. Klein, *Pacific Symposium on Biocomputing '97,* Singapore: World Scientific Press, January 1997.

Editor, with David J. States, Pankaj Agarwal, Terry Gaasterland and Randall Smith, *Proceedings of the Fourth International Conference on Intelligent Systems for Molecular Biology,* Menlo Park, CA: AAAI Press, July 1996

Editor, with Teri E. Klein, *Pacific Symposium on Biocomputing '96,* Singapore: World Scientific Press, January 1996.

Editor, with Christopher Rawlings, Dominic Clark, Russ Altman, Thomas Lengauer and Shoshana Wodak, *Proceedings of the Third International Conference on Intelligent Systems for Molecular Biology,* Menlo Park, CA: AAAI Press, July 1995

Editor, *Twenty-Seventh Annual Hawaii Internation Conference on System Sciences,* vol. 5: Biotechnology Computing, Los Alamitos, CA: IEEE Computer Society Press, Jan 1994

Editor, with Jude Shavlik & David Searls, *Proceedings of the First International Conference on Intelligent Systems for Molecular Biology,* Menlo Park, CA: AAAI Press, July 1993

Editor, with T. Mudge & V. Milutinovic*Proceedings of the Twenty-Sixth Annual Hawaii Internation Conference on System Sciences,* vol. 1: Computer Architecture and Biotechnology Computing, Los Alamitos, CA: IEEE Computer Society Press, Jan 1993

Other Publications:

Review of *Howard Aiken: Portrait of a Computer Pioneer* by I. Bernard Cohen, in *The New York Times Book Review*, September 12, 1999.

Creating a Professional Society for Bioinformatics - The International Society for Computational Biology (ISCB), with Christopher Rawlings. *Bioinformatics* 14: (6) 471-471 1998

Review of *Trapped in the Net* by Gene Rochlin, in *The New York Times Book Review*, Sept. 7, 1997.

Privacy Wrongs, with James Rule. *The Washington Monthly*, November 1996.

Review of *Computer: A History of the Information Machine* by Martin Campbell-Kelly and William Aspray. in *The New York Times Book Review,* Nov. 17, 1996

The State of Biotechnology Computing, 1994, *Proceedings of the Hawaiian International Conference on System Sciences* IEEE Computer Science Press, vol. 5, pp vi-viii, 1995

Introduction to the Special Issue on Molecular Biology Applications, with Jude Shavlik and David Searls. *Machine Learning*, 21: (1-2) 5-9 Oct-Nov 1995

Conference Report: The First International Conference on Intelligent Systems for Molecular Biology, with David

Searls and Jude Shavlik, *AI Magazine* 15(1):12-13, 1994

Public Image: Privacy in the Information Age. *Whole Earth Review*, 44:32-37, January 1985. Reprinted in *Social Issues Resource Services: Privacy, Volume 3,* 1986. Also reprinted in *The Borzoi College Reader,* eds. Charles Muscatine & Marlene Griffith, 7th edition, McGraw Hill, NY, 1992. Also reprinted in *Computers, Ethics and Social Values*, Deborah Johnson & Helen Nissenbaum, Prentice Hall, 1995.

A Report to ARPA on Twenty-First Century Intelligent Systems, with B. Grosz, R. Davis, R. Bajcsy, P. Bonisone, B. Bullock, S. Minton, T. Mitchell, R. Perrault, T. Lozano-Perez, M. Pollack, P. Rosenbloom, S. Shieber, H. Strobe and D. Weld, AAAI Press, Menlo Park, 1994

Review of Steven Levy's *Artificial Life* in *IEEE Spectrum* May 1993, 30(5):11-12.

Artificial Intelligence and Molecular Biology: Extended abstract of invited address, *Proceedings of the Tenth National Conference on Artificial Intelligence,* pp. 866-868, 1992, AAAI Press, Menlo Park, CA.

ARRIS: Searching for Drugs With AI Software *New Science: AI Research*, June 18, 1990, p. 1464

Industrial Applications of Machine Learning, New Science AI Industry Report, June 1989

Some Memory, but No Mind: A response to Smolensky's *On the Proper Treatment of Connectionism.* in *Behavioral and Brain Sciences*, 11(1), March 1988

AI Techniques: Analogical Reasoning. *New Science: AI Research*, June 20, 1988, p.709.

Review of Winograd and Flores *Understanding Computers and Cognition.* in *Technology Review*, July 1988

AI Techniques: Temporal Reasoning. *New Science: AI Research,* July 4, 1988, p. 719.

AI Attitudes and Techniques in Computer Supported Collaborative Work. *New Science: AI Research*, Aug. 15, 1988, p. 765

Review of Stewart Brand's *The Media Lab.* in *The New York Times Book Review,* Sept. 27, 1987, p. 38

Encapsulation and Expectation: A response to Fodor's *Modularity of Mind.* in *Behavioral and Brain Sciences,* 8(1): 29-30, 1985.

The Quest to Understand Thinking, with R. Schank. *Byte*, 10(4):143-155, April 1985.

Software Packages

COEV: A system for co-evolving learning agents and problem representations. Common Lisp and C code that implements a form of cultural co-evolution for synergistic multistrategy machine learning. A collection of diverse learning methods embodied as agents attempting to solve a particular problem evolve parameter settings via a genetic algorithm. The agents also generate partial solutions which compete with each other to be used by the learners, and in the process change the genetic fitness landscape of the learners. Patent pending, and licensed by several major US corporations.

Audio Knowledge Acquisition Tool, with Chuck McMath. A Macintosh application for the management of large amounts of audio protocol data. Distributed by the US National Technical Information Service; in use by knowledge engineers, psychologists, anthropologists and oral historians.

Amino Acid Representation Package. Common Lisp code for implementing a wide variety of representations for amino acids, including the novel Atoms-Orbitals-Hydrogens (AOH) representation. Used by machine learning researchers for protein structure prediction and other tasks.

AI & Molecular Biology Researchers Database. Database of names, contact information and research interests of more than 150 researchers worldwide. In 1995, the second most frequently accessed file in the European Molecular Biology Laboratory WAIS-server, widely used by students, academics and commercial organizations.

No longer maintained.

Prior and Active Research Support

NIH/CC Research Contract (Lawrence Hunter, Principle Investigator), 7/1/00-6/30/01, $100,000
Gene expression array analysis for critical care medicine studies. 20% effort
Performed gene expression array analysis and developed novel methods for the interpretation of data generated by NIH Clinical Center investigators in studies of sepsis and multiple organ failure.

1U01 AA13524-02 (Lawrence Hunter, Principle Investigator) NIH / NIAAA 9/1/01-8/31/06 $500,000
Neuroinformatics Core Facility for the Integrated Neuroscience Initiative on Alcoholism: 20% effort
The goal of this project is to develop a bioinformatics resource for a research consortium on alcoholism and neuroscience. The specific aims are: (i) integration of multiresolution neuroscience data, (ii) development of novel data mining tools to generate hypotheses on neuroadaptation to alcohol, and (iii) design and development of a web-based integrated computational analysis workbench for consortium investigators

Genetics Institute/Wyeth-Ayerst (Lawrence Hunter, Principle Investigator) 9/01/01-8/31/03, $113,650
Development of Biological Literature Text Mining Software (0% effort)
The purpose of this collaboration between the Expression Profiling Informatics ("EPI") group at Wyeth-Ayerst Research, and Professor Larry Hunter, Director of the Center for Computational Pharmacology at the University of Colorado School of Medicine, is to develop tools and software for automated literature mining. This support Funds a computational linguist research associate and related expenses.

1 R24 AA13162-01 (Boris Tabakoff, Principal Investigator) NIH / NIAAA 4/1/01-3/30/06, $999,562
Gene Expression Array Technology Center for Alcohol Research, 13.33% effort
The aim of this proposal is to establish a gene array technology core facility to serve as a national resource for alcohol research. The bioinformatics group will collaborate with NIAAA investigators in the analysis of expression array data and to develop a highly integrated database that includes gene expression profile data as well as genetic sequence and other data relevant to ethanol induced changes and ethanol susceptibility.

1M01 RR00051 (Robert Eckel, Principle Investigator) NIH / NCRR 4/01/02-3/31/07 $6,951,425
University of Colorado General Clinical Research Center 14.69% effort
The University of Colorado General Clinical Research Center has implemented an gene expression array facility for its users, and Dr. Hunter advises the bioinformatics director and his staff on appropriate analysis techniques for this novel and complex class of data.

5 P30 CA46934-15 (Paul Bunn, Principle Investigator) NIH / NCI 3/01/88-1/31/06
Cancer Center Support Grant, 23.51% effort
The University of Colorado Comprehensive Cancer Center (UCCC) is the only NCI- designed comprehensive Cancer Center in the Rocky Mountain region. Dr. Hunter is a member of the Biostatistics Core, and contributes to the design and analysis of gene expression array experiments and other bioinformatics issues that arise at the Center.

P01 HL68743 (Edward Abraham, Principle Investigator) NIH / NHLBI 9/01/02-8/31/07 $139,171

Heterogeneous neutrophil responses in acute lung injury, 10% effort
The overall hypothesis is that neutrophils produce heterogeneous responses to inflammatory stimuli. The Molecular Biology Core will perform microarray expression analysis on normal peripheral and BAL neutrophils, stimulated neutrophils and neutrophils from patients with acute lung injury. Dr. Hunter participates in gene expression array analysis for the Core.

P01 HL67671-01 (Robert Mason, Principle Investigator) NIH / NHLBI 7/01/01-6/30/04

SCOR: Pathobiology of Fibrotic Lung Disease, 10% effort
The overall purpose of this SCOR proposal is to investigate the role of the myofibroblast in idiopathic pulmonary fibrosis (IPF). Five projects investigate the source and regulation of TGF-beta production, especially the contribution of the ingestion of apoptotic cells and cell debris, the relationship of paracrine factors and mechanical factors on myofibroblast gene regulation, the role of survival factors for myofibroblasts such as IGF- I and myofibroblast apoptosis, interactions of myofibroblasts with alveolar epithelial cells, and finally regulation by interferon gamma (INF). Dr. Hunter performs informatics duties in the gene expression array core of the project.

Cystic Fibrosis Foundation, (David Rodman, Principle Investigator) 4/01/01-3/30/03, $500,000
Effects of Psuedomonas aeruginosa on Inflammatory Gene Expression, 3.48% effort
The aim of this proposal is to test the hypotheses that (1) Psuedomonas aeruginosa interacts with human airway epithelial cells and neutrophils to activate a pro-inflammatory patter of gene expression, (2) activation is more prominent in CF than non-CF epithelium and (3) specific gene products of P. aeruginosa can be identified as contributing to this aspect of bacterial virulence. The general experimental approach uses gene arrays, gene traps and proteomics. Dr. Hunter directs a bioinformatics group which will perform analyses of the array data.

RO1 HL ???? (Mark Geraci, Principle Investigator) NIH / NHLBI, 10/01/02-9/31/05, $500,000
Application of expression analysis to study disease pathogenesis. 10% effort
Create a shared microarray facility to support to NHLBI researchers for the incorporation of both cDNA and Affymetrix expression arrays into their research endeavors. Specific aims are to perform developmental projects for maximizing RNA amplification techniques and utilizing reference standards and strategies to develop algorithms for direct comparison of data from cDNA arrays and Affymetrix arrays; and to develop and implement novel bioinformatic approaches to expression data analysis, including "scripted" internet-based analysis for NHLBI researchers. Dr. Hunter directs the bioinformatics effort.

1 R01 DE 15191-01 (Richard Spritz, Principle Investigator) NIH / NIDCR, 2/01/03-1/31/07, $250,000
Gene Discovery for Craniofacial Disorders 5% effort
The goal of the proposal is to identify the genes, pathways, and genetic networks that are involved in craniofacial development and thus represent targets for genetic and non-genetic determinants of non-syndromic cleft lip and/or palate. We plan a careful microarray study of gene expression profiles in the developing face of the mouse. Dr. Hunter will apply state of the art bioinformatics tools to analyze and interpret the data.

Pending Research Support

NIH 1R01 LM008111-01 (Lawrence Hunter, Principle Investigator), 12/1/03-11/30/06 $499,000
Technology Development for a Molecular Biology Knowledge-base 15% effort
The goal of this proposal is to demonstrate that database integration and natural language information extraction technology are adequate to produce in automated fashion a broad, deep knowledge-base of molecular biology.

1R37 HD19547-19 (Margaret Neville, Principle Investigator) 7/1/03-6/30/08, $250,000
Physiological factors affecting Human Lactation, 5% effort
Renewal of Dr. Neville's grant for studies of milk secretion and its regulation. Dr. Hunter would be added to oversee bioinformatic analysis of gene expression array studies.

Selected Lectures and Presentations:

The Era of Biognostic Machines, keynote address to Association for Computing Machinery Special Interest Group on Applied Computing (ACM-SAC) conference., 2003

Proteomic Bioinformatics, Center for Computational Pharmacology mini-symposium, 2003

Biognostic Machines for Cognitive Disability, invited address, Coleman Institute annual meeting, 2002

Bioinformatics and Human Health, UCHSC Chancellor's Luncheon Address, 2002

Data Mining for High Throughput Biomedicine, keynote address to the Research Society on Alcoholism conference, Denver, Colorado, June 2000

Edgar: Extraction of Drugs, Genes and Relations from the Biomedical Literature, Pacific Symposium on Biocomputing, January, 2000

The Role of Machine Learning and Natural Language Processing in Contemporary Drug Discovery, Pharmacology Grand Rounds, University of Colorado School of Medicine, October, 1999

Inductive Modeling: Power and Pitfalls, keynote address to MODEL-IT conference, Waginengen, the Netherlands, November 1998

Coevolution of Symbol Systems and Behavior, lecture and workshop, Simulations of Adaptive Behavior conference, Zurich, Switzerland, August 1998.

Machine Learning for Drug Discovery, invited address, SmithKline Beecham Data Mining Days, November 1997.

Computer Science : Biology :: Mathematics : Physics, MIT Media lab, April 1997

The Role of Computation in Cognitive Science, Krasnow Institute for Advanced Study of Cognition Seminar Series, November, 1996.

Coevolution Learning: Syngerstic Evolution of Learning Agents and Problem Representations, Multistrategy Learning Workshop, June, 1996.

AI Models for Biology, and Biological Models for AI, Keynote address, Second International Conference on Intelligent Systems for Molecular Biology, July 1995.

Computers, Modelling , and Theorretical Biology, Invited address to the Keystone Center Scientist to Scientist Colloquium, August, 1994

The National Library of Medicine on the Internet: A Digital Library for Biomedicine. Invited address to the Computers and Chemistry Division of the American Chemical Society conference, Aug 1994

Planning to Discover in Molecular Biology, MIT AI Lab Revolving Seminar Series, April 1994

Molecular Biology for the Computer Scientist, Full day tutorial at the Hawaiian International Conference on System Sciences, January 1993. Repeated Jan 1994.

AI & Molecular Biology, Plenary address, National Conference on Artificial Intelligence, San Jose, CA, July 1992.

Megaclustering of Unsegmented Datastreams and Applications to Molecular Biology, Johns Hopkins Applied Physics Laboratory distinguished lecture series, October 1992.

Electronic Facilitation of Scientific Communication, Panel organizer and speaker, International Conference on the Biomatrix, George Mason University, July 1990

Knowledge Acquisition Planning for Inference from Large Datasets, Keynote address, 1990 Conference on AI Systems in Government, Washington, DC, May 1990

Machine Learning: Ready for Industrial Application, Invited address to Third Annual Artificial Intelligence Forum, Sanibel Island, FL, February, 1989

Artificial Neural Networks as Theories of Mind. International Neural Network Society, Boston MA, September, 1988

Machine Learning for Molecular Biology. Invited address to the Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, June 1988

Indexing and Recognition. AI/BioMed: The First International Conference on Artificial Intelligence and its Impacts in Biology and Medicine, Montpellier, France, September 1986

Computers and Privacy. Guest lecture in Constitutional Law, University of Connecticut at Hartford Law School, Dec., 1985.